

# Weekly Report

April 30, 2017

## 1 Work

本周，在linux上运行了EFANNA和LargeVis，并且了解了tsne，bh-tsne和largevis，对于高维数据投影有了更深入的了解。同时，本周审阅了论文《PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models》。

## 2 Paper Reading

### 2.1 Visualizing High-Dimensional Data Using t-SNE

tsne投影分两步，第一步是构建投影前后的概率分布（比如，两点的距离在所有距离和中所占的比例，距离越小比例越大），第二步是根据两个概率分布求一个熵作为目标函数，两个概率分布越接近，目标函数越小。其中通过使用不同的概率分布函数，使得投影后，在原空间比较接近的点，在投影空间需要更接近才能获得相同的概率，而在原空间本身比较分开的点，需要更远才能获得相同的概率。这样的结果就会使得，相同的数据比较接近，不同的数据比较远。

### 2.2 Accelerating t-SNE using Tree-Based Algorithms

tsne的作者在2014年发布的论文，加速了tsne。第一步的想法是在计算概率分布的过程中，距离非常远的点，对于总体的概率分布的影响不大，可以忽略，所以计算概率分布只要基于每个点的k个最近邻即可。第二步是说，在优化过程中，不需要计算点和一个较远区域所有点的距离，因为这些距离都比较接近，可以只计算一个来近似。

## 2.3 PhenoLines: Phenotype Comparison Visualizations for Disease Subtyping via Topic Models

本文使用了主题模型自闭症谱系障碍的儿童病例进行了聚类观察。每个年龄阶段的儿童症状都被一起聚类成几个主题，每个症状会有一定的对于主题的概率分布。作者吧症状按照医学中的分类组织成一个图结果方便用户进行有层次的分析。



Figure 1: 主要可视化的功能包括四块，radial hierarchy，scatterplot，rank-ordered list，stacked timelines

## 2.4 Blockwise Human Brain Network Visual Comparison Using NodeTrix Representation

本文主要是可视化了人脑神经网络之间的关系，通过聚类等方法减少可视化遮挡，然后使用NodeTrix探索区域之间的联系。

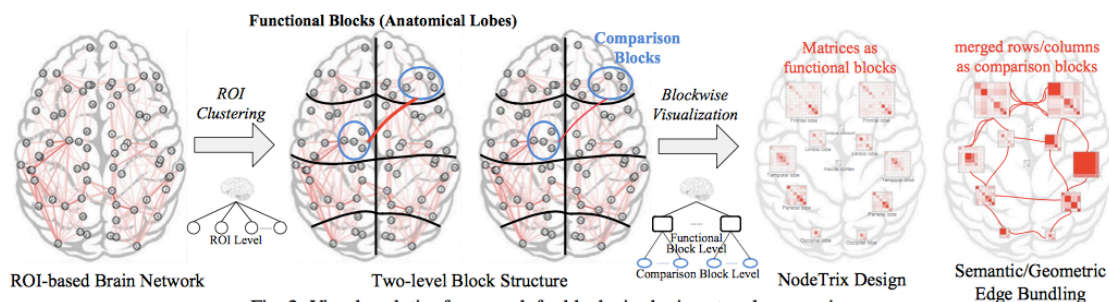


Fig. 2: Visual analytics framework for blockwise brain network comparison.

Figure 2:

## 2.5 EventAction: Visual Analytics for Temporal Event Sequence Recommendation

推荐系统是我们在网上经常可以接触的，然而从未有人研究过对于时序事件的推荐系统。基于相似事件的判别，系统会推荐根据和你有类似经历的人的事件。